REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
29 Dec 97

3. REPORT TYPE AND DATES COVERED
Final Progress

4. TITLE AND SUBTITLE
Modeling and Inversion of Shallow Data Including Nongeometrical Waves

5. FUNDING NUMBERS
DAAH04-95-1-0173

6. AUTHOR(S)
Dr. John A. Scales, Dr. W.C. Navidi, Mr. Alberto Villarreal

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Center for Wave Phenomena
Colorado School of Mines
Golden, CO 80401

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 32891.11-MS

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

19980521 124

13. ABSTRACT *(Maximum 200 words)*
Solving a geophysical inverse problem requires making inferences about the earth from data. Since one always has only a finite number of (uncertain) data and since the models used to describe the earth are infinite dimensional (i.e., functions of space), it follows that if there are any models at all that fit the data, there will likely be many of them. Thus, finding a single model that fits the data is of limited value without a quantitative assessment of its uncertainty. During the course of this project we have developed novel theoretical and computational strategies for making statistically rigorous inferences about the earth's near-surface from full-waveform reflection and borehole seismic data. Our approach allows us to assimilate information at vastly different length scales and to take advantage of all the information in the seismic waveforms, as well as quantifying uncertainties in the data due to noise and theoretical errors. We have demonstrated the efficiency and utility of this approach on field data and have produced computer codes (using freely available compilers and message passing libraries) which perform in a scalable, distributed-parallel fashion on heterogeneous networks of workstations or shared-memory multiprocessors.

14. SUBJECT TERMS

15. NUMBER IF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

FINAL REPORT

FOR THE ARMY RESEARCH OFFICE

1. ARO PROPOSAL NUMBER: 32891–GS.

2. PERIOD COVERED BY REPORT: 1 April, 1995 – 31 December, 1997.

3. TITLE OF PROPOSAL: Modeling and Inversion of Shallow Seismic Data Including Nongeometrical Waves.

4. CONTRACT OR GRANT NUMBER: DAAH04-95-1-0173.

5. NAME OF INSTITUTION: Center for Wave Phenomena, Colorado School of Mines.

6. AUTHORS OF THE REPORT: Dr. John A. Scales, Dr. William Navidi, Mr. Alberto Villarreal

7. LIST OF MANUSCRIPTS SUBMITTED OR PUBLISHED UNDER ARO SPONSORSHIP DURING THIS PERIOD, INCLUDING JOURNAL REFERENCES:

1. Tsvankin, I., 1995, Seismic wavefields in layered isotropic media: Samizdat Press (available via the WWW at `http://landau.mines.edu/~samizdat`.

2. Deng, H., W. Gouveia, and J.A. Scales, *The CWP object-oriented optimization library*, The Leading Edge, **15**, 365–369, 1996.

3. Deng, H., W. Gouveia, and J.A. Scales, *An object-oriented toolbox for studying optimization problems*, 320–330, in Inverse Methods, edited by B.H. Jacobson, K. Mosegaard, and P. Sibani, Springer-Verlag, 1996.

4. Scales, J.A., *Uncertainties in seismic inverse calculations*, 79–97, in Inverse Methods, edited by B.H. Jacobson, K. Mosegaard, and P. Sibani, Springer-Verlag, 1996.

5. Villarreal, A. and J.A. Scales, *Distributed 3D finite difference modeling via domain decomposition*, Computers in Physics, **11**, 388–399, 1997.

6. Gouveia, W. and J.A. Scales, *Resolution of seismic waveform inversion: Bayes versus Occam*, Inverse Problems, **13**, 323–349, 1997.

7. Gouveia, W. and J.A. Scales, *Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis*, in press Journal of Geophysical Research, 1997.

8. Van Wijk, K., J.A. Scales, W. Navidi and K. Roy-Chowdhury, *Automatic estimation of data uncertainties for least squares optimization*, submitted to JGR, 1997.

8. SCIENTIFIC PERSONNEL SUPPORTED BY THIS PROJECT AND DEGREES AWARDED DURING THIS REPORTING PERIOD: Dr. John A. Scales, Dr. Ilya D. Tsvankin, Wences Gouveia, Alejandro Murillo, Alberto Villarreal, Dr. Konstantin Osypov, and Dr. William Navidi. PhD awarded to Wences Gouveia.

PROJECT GOALS

The purpose of this project was to use full-waveform seismic inversion methods, including non-geometrical waves not traditionally treated in seismic imaging algorithms, to produce high resolution images of the near surface of the earth along with rigorous estimates of the uncertainty of these images. Our strategy is to

- Incorporate sophisticated geologic *a priori* information and estimates of data uncertainties into a rigorous statistical framework.
- Make quantitative estimates of the uncertainties in the estimated model parameters. I.e., put "error-bars" on all computed results.
- Treat the full elastic wavefield without kinematical approximation.
- Use state-of-the-art optimization methods.
- Make parallel, distributed-memory implementations of the core numerical algorithms, so as to achieve a low-cost scalability and efficiency.

This approach is novel in that it aims not only to use the maximum information available in the seismic waveforms, but to integrate this waveform data with sophisticated geologic, geophysical and petrophysical information in a statistical framework that allows us to assign quantitative uncertainty estimates to all computed parameters. The technical barriers that motivated this study were primarily 1) the inherent limitations of kinematical inversion methods in complex near-surface settings, and 2) the inability of existing inversion methods to deal in a rigorous fashion with diverse forms of geological and geophysical information available.

TECHNICAL OVERVIEW

Our strategy for performing the inversion was to carefully estimate all the significant uncertainties in the data. In our field data studies we estimated the errors not only associated with ambient noise, but also errors in the data processing (such as residual statics corrections), theoretical errors associated with discretized models used, and finally errors associated with the unknown scaling between field and synthetic data. By knowing the uncertainties in different, independent data sets these data sets can be combined into a single inverse calculation without arbitrary scaling factors.

But for any reasonably rich parameterization of the subsurface, *a priori* unreasonable models will fit the data too. Therefore we have adopted a Bayesian strategy to assign prior probability to earth models derived from, in our case, *in-situ* petrophysical measurements—well logs, for instance. To rigorously assess the significance of this combination of information we compute uncertainty estimates based both on the prior information and on the combination of prior information and the seismic data. This comparison gives us a quantitative measure of resolution that takes into account all of the information.

To make all this efficient enough to be applicable to large-scale seismic data sets we have developed fast, parallel waveform modeling and optimization codes. These codes, which are freely available from the Center for Wave Phenomena WWW site: `http://www.cwp.mines.edu` are designed to be run efficiently on a network of workstations or on high-performance shared-memory supercomputers.

MAIN RESULTS

Prior Information.—

We have developed new techniques for rigorously integrating diverse sources of geological and geophysical information into near-surface seismic inverse problems (cf. Gouveia and Scales (1997a) and Gouveia and Scales (1997b). This information could include *in-situ* petrophysical measurements, laboratory measurements and geological observations. Since we actually estimate the uncertainties of all the measurements, which are independent, it is straightforward to assimilate these diverse measurements in a Bayesian framework.

Once we have the prior information, then solving the inverse problem amounts to constructing a final (or posterior) probability density on the space of models; statistical inferences can then be extracted from from this posterior by integration. Further, particular models can be found by optimizing the posterior. In Scales (1996) it is shown that the posterior can be deduced from the prior information by a straightforward application of Bayes' theorem. This approach differs fundamentally from the common treatment of Bayesian inversion in which the posterior is a conditional probability conditioned on the data. In Gouveia and Scales (1997b) we show a field data case study of the application of this methodology to multiple well logs and surface seismic data.

Waveform Modeling.—

In order to make multi-offset seismic inversion feasible on workstations we devoted a substantial effort to producing network-parallel implementations of all the modeling code. The papers by Gouveia and Scales (1997b) and Villarreal and Scales (1997) show two different approaches. In the former we treated earth models as being layered. This leads to substantial theoretical simplification. The full anelastic equations of motion can be solved essentially analytically by the reflectivity method. This is a frequency-domain approach and since each frequency component can be computed independently, the method parallelizes easily using a master/slave approach. The master processor distributes blocks of frequencies over the network (according to processor power and load) and then reassembles the final time-domain result, as illustrated in Figure 1. Further, the Frechet derivatives needed to perform local optimization can be computed analytically as well.

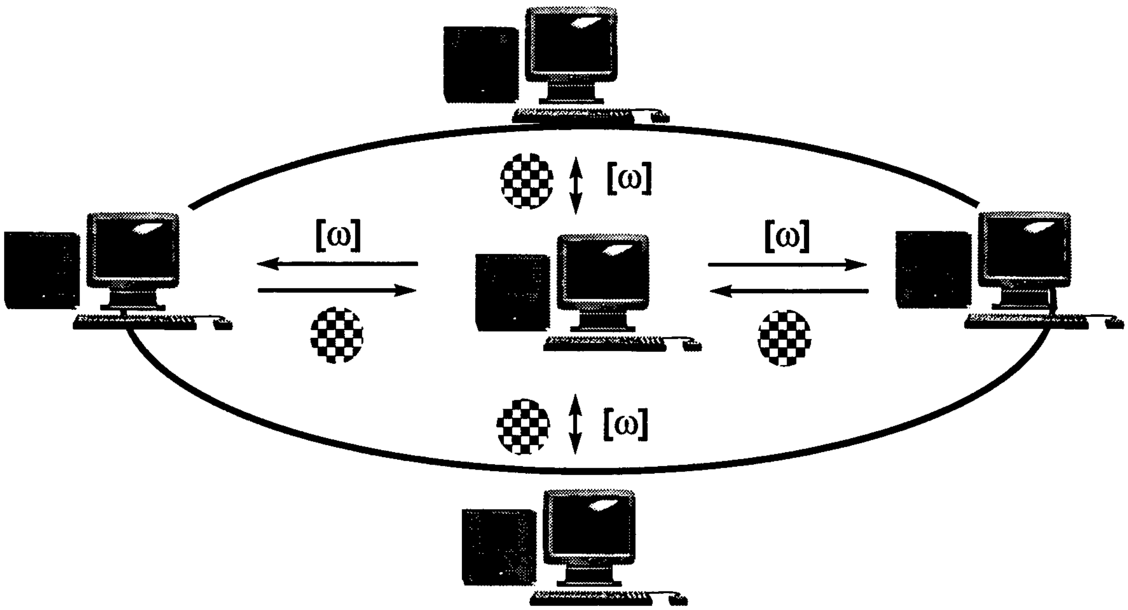

FIG. 1. The distributed-memory model used for the elastic reflectivity modeling code. The master processor distributes blocks of frequencies to the slave processors, when then reassemble the full seismograms.

On the other hand, to get beyond layered earth models requires the use of finite element or finite difference methods. These methods are more difficult to parallelize in distributed fashion. However in Villarreal and Scales (1997) we describe a domain-decomposition approach that allows us to solve 3D acoustic and elastic modeling problems on our network of PCs. Further, by taking advantage of freely available message passing libraries such as PVM, both the reflectivity and the finite difference codes can be run without change on large shared memory machines such as the SGI Power Challenge or the IBM SP2.

Optimization.—

The computational core of most inversion algorithms is an optimization calculation, minimizing or maximizing some function subject to constraints and penalties. We have developed an extensive library of efficient optimization algorithms (Deng *et al.* 1996a; Deng *et al.* 1996b) that lets us tailor the optimization to suit the problem at hand. This library is called COOOL (for CWP Object-Oriented Optimization Library) and is illustrated in Figure 2 and Figure 3. These codes, which were designated by PC Computing magazine as one of the "1001 Best Internet Downloads" (July, 1997), can be downloaded from the site:

`ftp://ftp.cwp.mines.edu/pub/cwpcodes/coool`

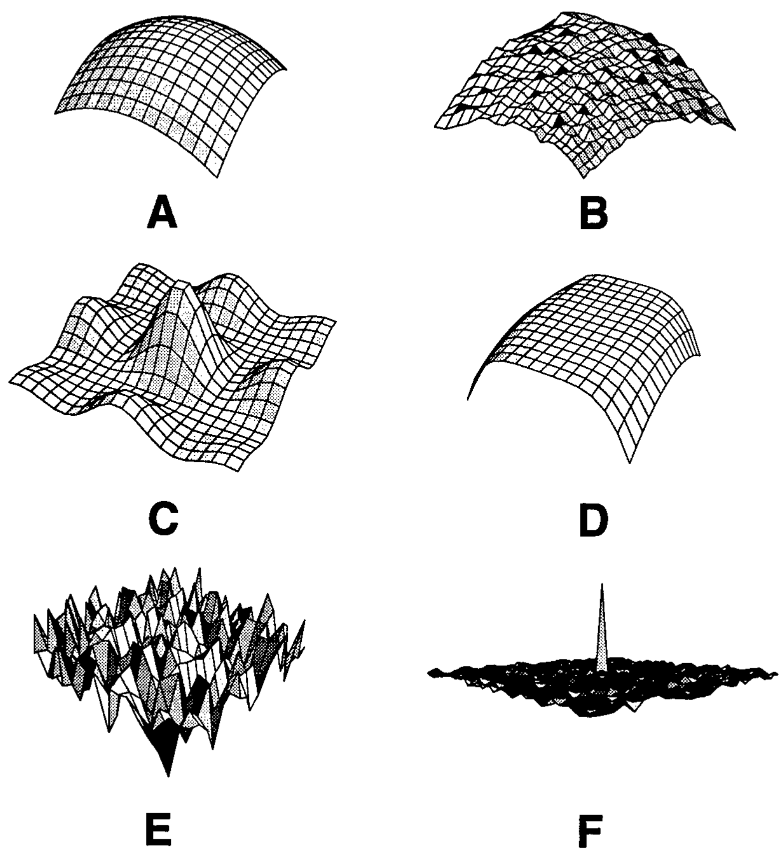

FIG. 2. The rich diversity of functions found in geophysical inverse problems. To be able to handle all of these kinds of functions we need a rich library of optimization routines, such as provided by the COOOL library. A: unimodal, a single extremum. B: essentially unimodal but with parasitic local extrema. C: fundamentally multimodal, small number of local extrema. D: significant null-space effects. E: fundamentally multimodal, huge number of local extrema. F: lacking any useful structure, brute force probably required. To aid in the visualization, all the function examples shown are are functions of only two dimensions. In practice, we are faced with functions of hundreds or thousands of unknowns.

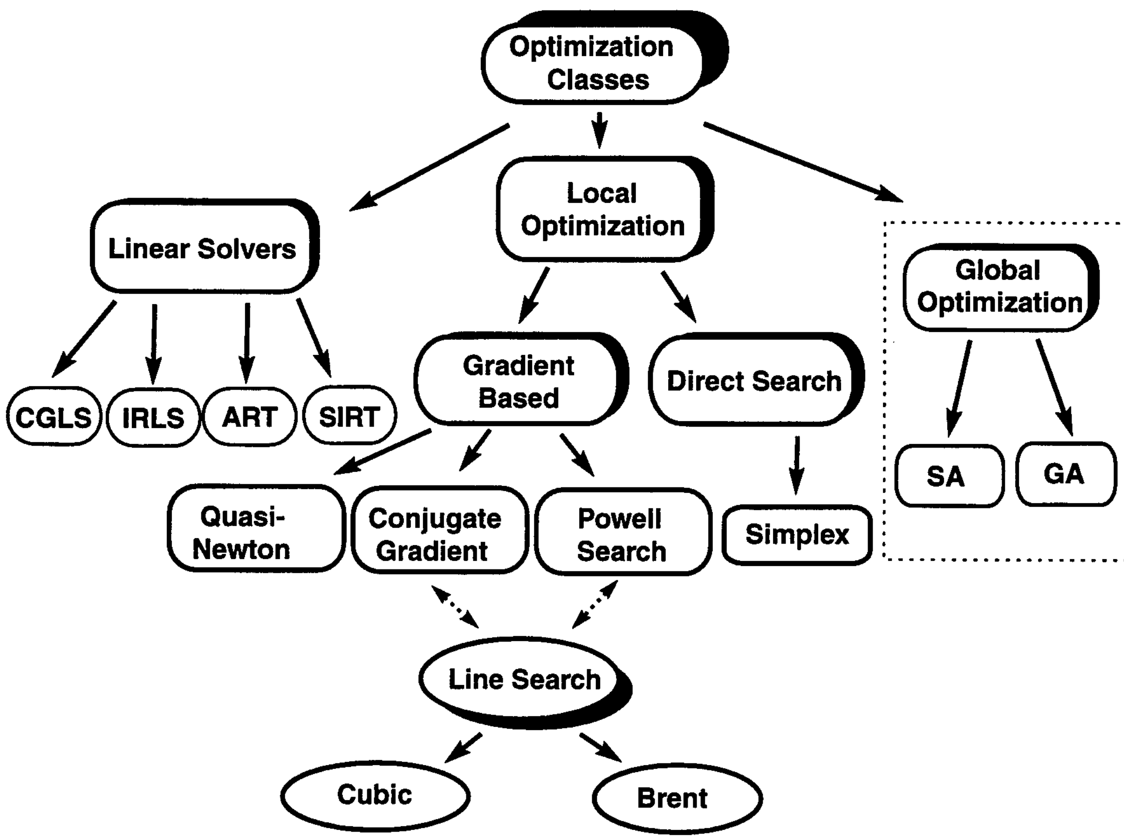

FIG. 3. An overview of the COOOL library of C++ optimization codes.

A CASE STUDY

In addition to the new theory and software developed in the course of this project, we performed a full-scale application the methods to a subset of a large 3D reflection seismic survey The surface seismic data and well logs we used were provided by T. Davis of the Colorado School of Mines Reservoir Characterization Project. The seismic data are a small subset of the vertical component data extracted from a nine-component survey acquired at the Sorrento Basin (near the Las Animas Arch, southeast Colorado). For this study we integrated surface seismic data (a sample is shown in Figure 4) and well logs (Figure 5).

Figure 6 shows a concrete example of the sort of ambiguity that is ubiquitous in geophysical inverse theory. It shows two layered earth models (P-wave impedance, S-wave impedance and density as a function of depth) computed from the vertical component Sorrento data. Below these models are shown the recorded data and the response for the corresponding models. The recorded and computed traces are alternated in both plots in order to show clearly the extent of the data fit. A careful analysis of the uncertainties in these data show that both models fit the data in a rigorous statistical sense. The elastic wavelengths (roughly 100m or so) are much larger than the discretization level used in the calculation, so clearly much of the structure seen in the model on the left is not required to fit the surface data.

The model labeled Occam is the smoothest layered model that fits the surface data; it therefore represents the broadest average of the earth that is capable of fitting the surface data. The model on the left (labeled Bayes) not only fits the surface data (as is evident from the bottom left plot) but it also fits a nearby well log. If we were to ignore the well log data, then this figure is a clear illustration of the large ambiguity in surface seismic data. By taking into account *a priori* information or other data sets (e.g., the well log) we can substantially reduce the uncertainty in our computed models. Thus the key step in performing *data fusion* must be to rigorously estimate the uncertainties in the data.

Once we have integrated the data and prior information using Bayes Theorem, it is a matter of extracting rigorous inferences from the posterior probability distribution. To make the calculation even more efficient, we make a Gaussian approximation about the peak of the posterior (the so-called Maximum *A Posteriori* model). The width of this distribution about the MAP model, measured by the *a posteriori* covariance matrix, gives a quantitative measure of the resolution of the calculation. A comparison of the prior posterior uncertainties is a direct measure of resolution. Figure 7 shows such a comparison for two particular shot records at the extreme ends of data quality. This figure, or ones like it, is the ultimate goal of our calculation–a concrete measure of the resolution of the data sets that takes into account all of the significant uncertainties in the calculation and allows one to fairly estimate the risk of various interpretations of the data. Finally we show in Figure 8 the MAP model bracketed by plus or minus one-standard deviation error bars. A coverage of two standard deviations represents approximately 70% of the area under a normal distri-

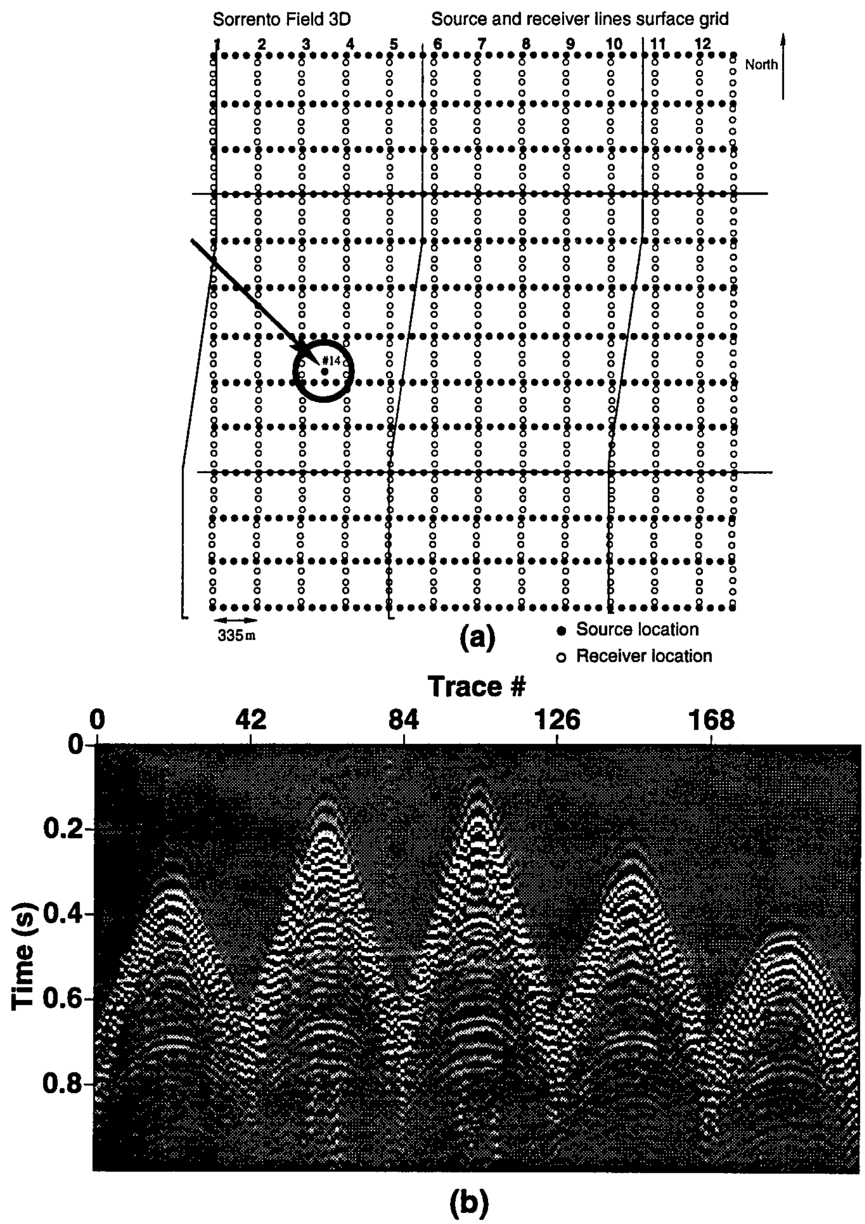

FIG. 4. (a) Intended source and receiver positions for the 3D Sorrento Survey. The shot gathers within the box will be denoted shots 1 to 5, starting from the leftmost one. The arrow indicates the location of the well MULL 14. (b) Lines 1, 2, 4, 5 and 6 of shot gather 1. Receiver-group spacing is 67 m, receiver-line spacing is 335 m and time sampling interval is 4 ms.

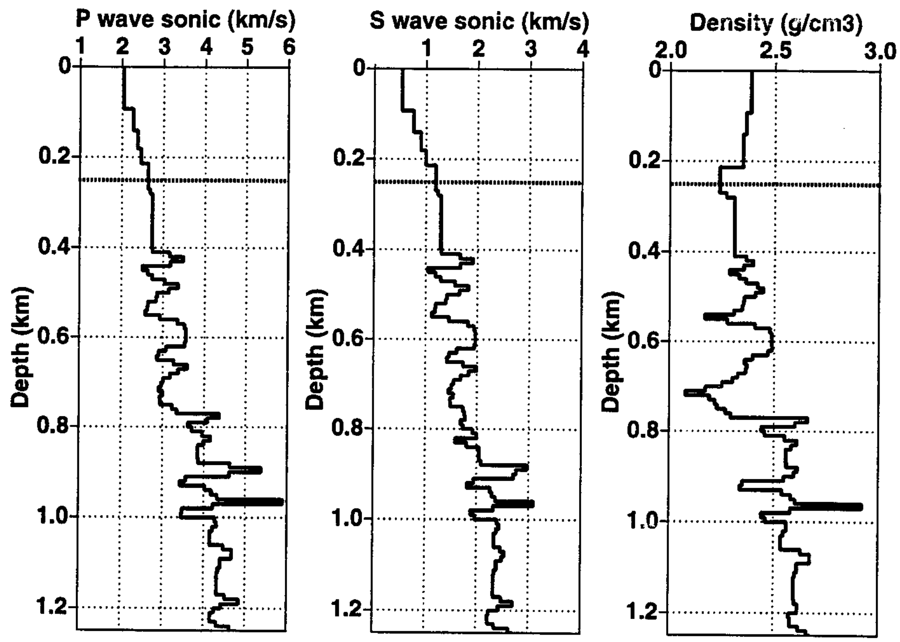

FIG. 5. Well logs acquired at Mull 14 after median filtering and blocking. The target depth interval for the inversion is 1 km thick and goes from 0.25 km (dashed line) to 1.25 km. The discretization interval is 10 m.

bution. So we can interpret this result as showing a region within which we the true value of the parameter lies at the 70% confidence level. If we need need a higher level of confidence we must use larger error bars. For instance plus or minus two standard deviations would give us 95% confidence.

CONCLUSIONS

Someone once said that the difference between theory and practice is larger in practice than in theory. Our goal has been to lessen the difference. Our field data case study (Gouveia and Scales (1997b) was described by an editor of the *Journal of Geophysical Research* as containing "a statistical treatment of data uncertainties and a priori information that is far more ambitious than seen hitherto ... the paper represents one of the most careful studies I have seen within this field." One of our papers on the theoretical aspects of the method (Gouveia and Scales, 1997a) was described by the reviewer as "the missing piece in the debate opposing Bayesians and Occamists ... Fundamental concepts in parameter estimation such as resolution and bias are revisited, which brings much insight to the reader on both philosophies. This paper is thus an essential piece of work to add to inverse problem theory applied to seismic exploration." Our latest work (Van Wijk et al.(1997)) develops new algorithms that will make it easy for non-experts to estimate the uncertainties of their data automatically. By demonstrating the possibility to apply rigorous statistical methods to large-scale full-waveform seismic inversion, we believe this work represents a significant contribution both to the study of geophysical inverse problems and

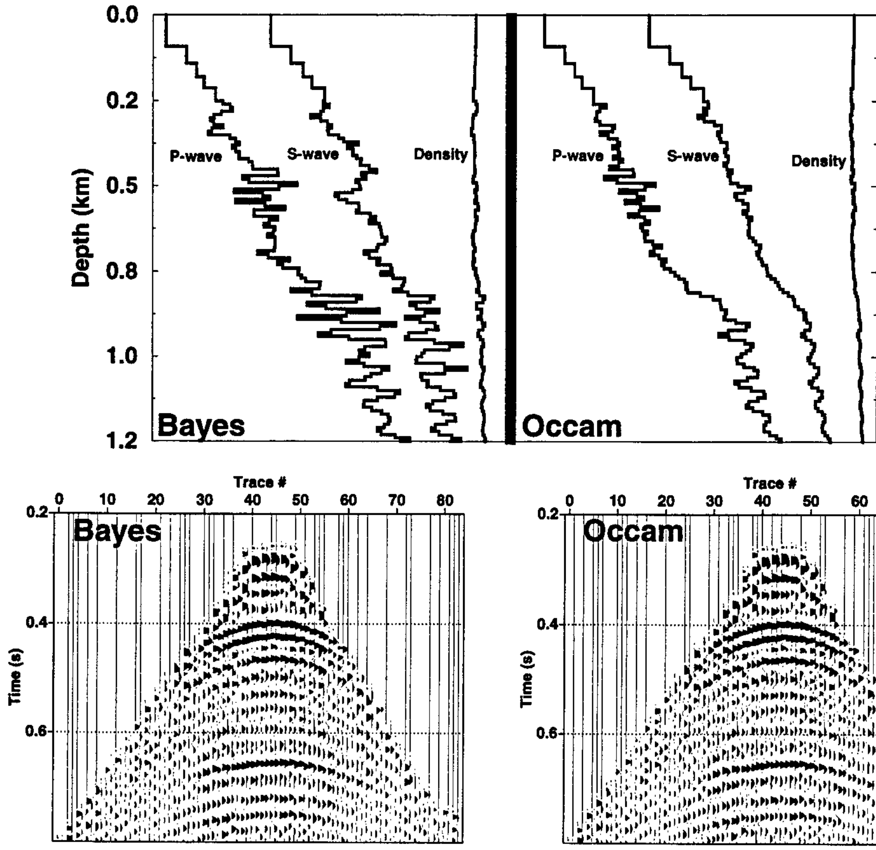

FIG. 6. Layered earth models (the three traces correspond respectively to P-wave impedance, S-wave impedance and density as a function of depth) computed by two different methods of inverting a single, vertical component common source record. The model on the right ("Occam") is the smoothest model that fits the data; it therefore represents the broadest average of the earth that is capable of fitting the surface data. The model on the left not only fits the surface data, but is statistically consistent with a nearby well log. Below, observed and response traces are alternated.

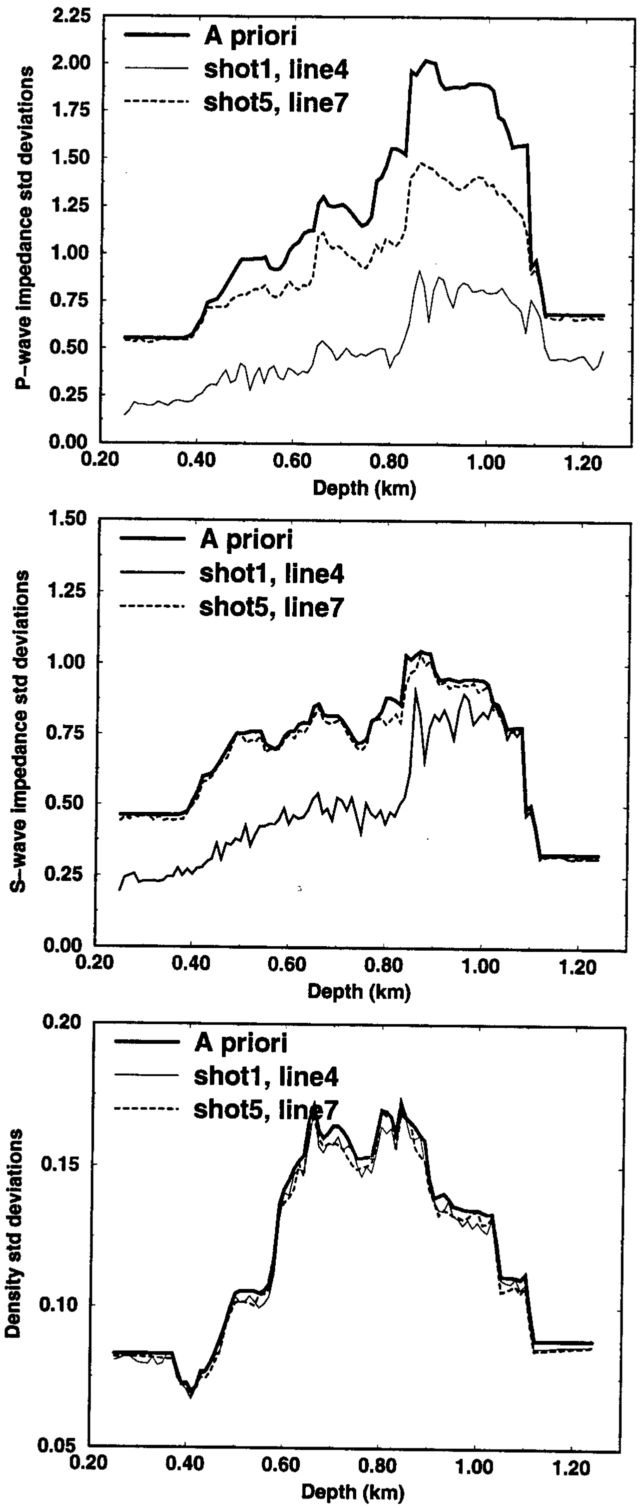

FIG. 7. A comparison of prior and posterior standard deviations for two particular shot records, one having good signal/noise and one poor. We see that in both cases there is no resolution of density at all. P-wave impedances are relatively well resolved, S-wave impedance less so. (We used vertical component seismograms.) These uncertainties (strictly the full covariance matrix) take into account both data fit and the prior well log information. Together with the peak of the posterior distribution (shown in the upper left part of Figure 6) they constitute the solution of the inverse problem within the Gaussian approximation.

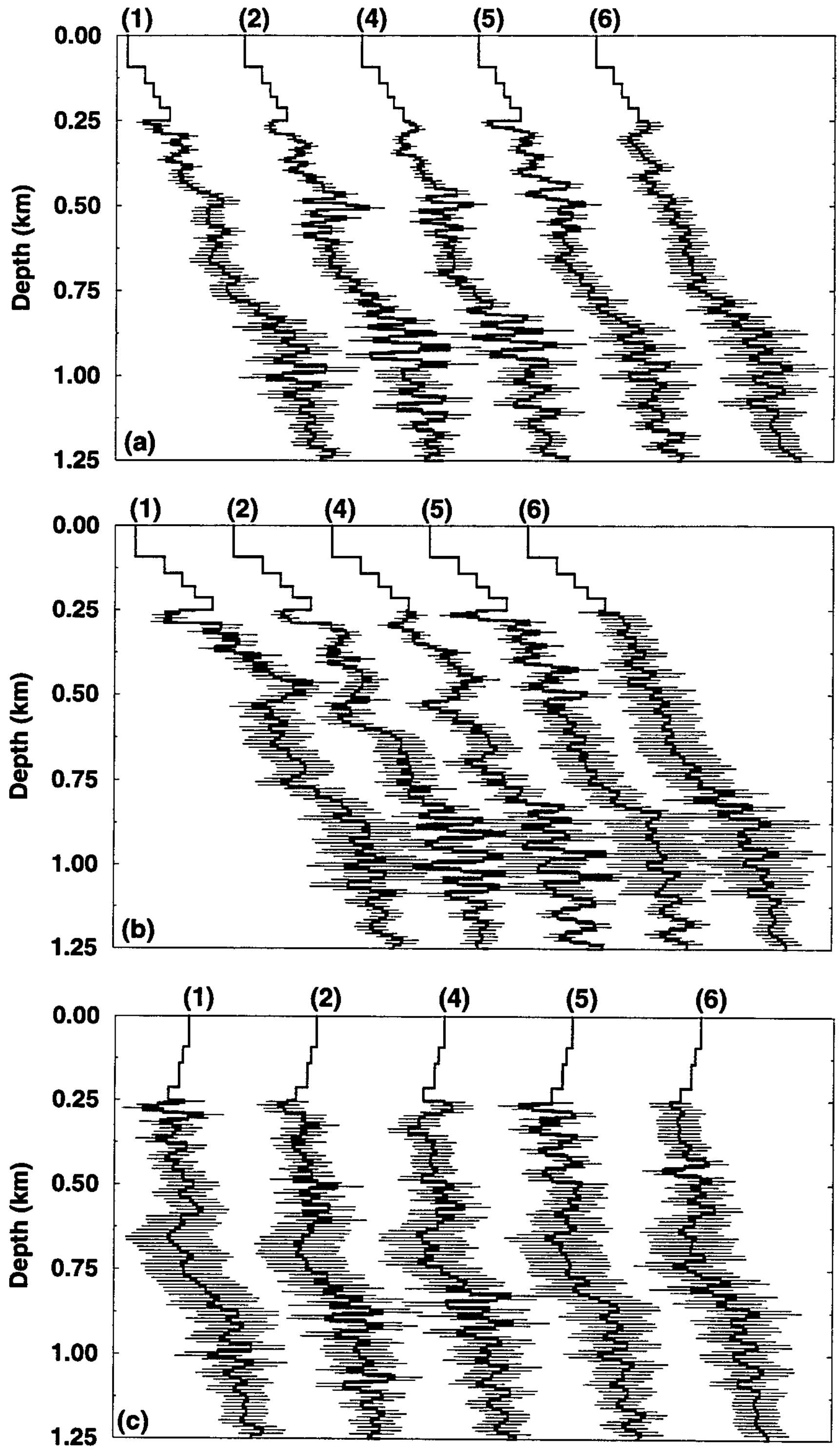

FIG. 8. MAP models derived from shot gather 1 (Figure 4): (a) P-wave impedance profiles. (b) S-wave impedance profiles. (c) Density profiles. The numbers at the top of each profile are associated with the line numbers. The error bars are ± unit standard deviations derived from the *a posteriori* covariance matrix.

to the very practical problem of quantitatively estimating the risk associated with imaging the near surface.